



IJCNLP 2011
Proceedings of
the Workshop on
Advances in Text Input Methods
(WTIM 2011)

November 13, 2011
Shangri-La Hotel
Chiang Mai, Thailand



IJCNLP 2011

**Proceedings of
the Workshop on Advances in Text Input Methods
(WTIM 2011)**

November 13, 2011
Chiang Mai, Thailand

We wish to thank our sponsors

Gold Sponsors



www.google.com



www.baidu.com



[The Office of Naval Research \(ONR\)](#)



[The Asian Office of Aerospace Research and Development \(AOARD\)](#)



[Department of Systems Engineering and Engineering Management, The Chinese University of Hong Kong](#)

Silver Sponsors



[Microsoft Corporation](#)

Bronze Sponsors



[Chinese and Oriental Languages Information Processing Society \(COLIPS\)](#)

Supporter



[Thailand Convention and Exhibition Bureau \(TCEB\)](#)

We wish to thank our sponsors

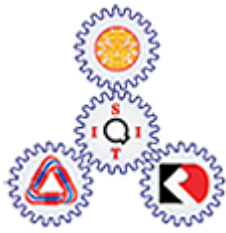
Organizers



[Asian Federation of Natural Language Processing \(AFNLP\)](#)



[National Electronics and Computer Technology Center \(NECTEC\), Thailand](#)



[Sirindhorn International Institute of Technology \(SIIT\), Thailand](#)



[Rajamangala University of Technology Lanna \(RMUTL\), Thailand](#)



[Maejo University, Thailand](#)



[Chiang Mai University \(CMU\), Thailand](#)

©2011 Asian Federation of Natural Language Processing

Preface

Welcome to the IJCNLP Workshop on Advances in Text Input Methods (WTIM 2011)!

Methods of text input have entered a new era. The number of people who have access to computers and mobile devices is skyrocketing in regions where people do not have a convenient method of inputting their native language. It has also become commonplace to input text not through a keyboard but through different modes such as voice and handwriting recognition. Even when people input text using a keyboard, it is done differently from only a few years ago – adaptive software keyboards, word auto-completion and prediction, and spell correction are just a few examples of such recent changes in text input experience. The changes are global and ubiquitous: users are no longer willing to input text without the help of new generation input methods regardless of language, device or situation.

The challenges in text input have many underlying NLP problems in common. For example, a high quality dictionary is called for, but it is far from obvious how to construct and maintain one. A dictionary also needs to be stored in some data structure, whose optimal design may depend upon the usage. Prediction and spell correction features can be very annoying if they are not smart enough. For many applications, user input can be very noisy (imagine voice recognition or typing on a small screen), so the input methods must be robust against such noise. We expect input methods to learn from the history of text input, but we are yet to see such an intelligent system. Finally, there is no standard data set or evaluation metric, which is necessary for quantitative analysis of user input experience.

The goal of this workshop is to bring together the researchers and developers of text input technologies around the world, and share their innovations, research findings and issues across different applications, devices, modes and languages. This volume contains contributions on diverse aspects of text input methods research on a variety of languages. We hope that the workshop serves as a starting point for deepening our understanding of the field as a whole, and for facilitating further innovations in user text input experience.

Hideto Kazawa, Hisami Suzuki and Taku Kudo
Organizers, WTIM 2011

Organizers:

Hideto Kazawa (Google, Japan)
Hisami Suzuki (Microsoft Research, USA)
Taku Kudo (Google, Japan)

Program Committee:

Achraf Chalabi (Cairo Microsoft Innovation Center, Egypt)
Frank Yung-Fong Tang (Google, USA)
Haifeng Wang (Baidu, China)
Hiroyuki Tokunaga (Preferred Infrastructure Inc, Japan)
Jianfeng Gao (Microsoft Research, USA)
Kalika Bali (Microsoft Research India)
Kazuma Takaoka (JustSystems, Japan)
Kumiko Tanaka-Ishii (University of Tokyo, Japan)
Mamoru Komachi (Nara Institute of Science and Technology, Japan)
Mike Schuster (Google, USA)
Monojit Choudhury (Microsoft Research India)
Shinsuke Mori (Kyoto University, Japan)
Thanaruk Theeramunkong (SIIT, Thailand)
Virach Sornlertlamvanich (NECTEC, Thailand)
Yoh Okuno (Yahoo! Japan)

Co-sponsor (Keynote speech):

Nara Institute of Science and Technology, Japan

Keynote speaker:

Toshiyuki Masui (Keio University, Japan)

Workshop Homepage:

<https://sites.google.com/site/wtim2011/>

Table of Contents

<i>Challenges in Designing Input Method Editors for Indian Languages: The Role of Word-Origin and Context</i>	
Umair Z. Ahmed, Kalika Bali, Monojit Choudhury and Sowmya VB	1
<i>Discriminative Method for Japanese Kana-Kanji Input Method</i>	
Hiroyuki Tokunaga, Daisuke Okanohara and Shinsuke Mori	10
<i>Efficient dictionary and language model compression for input method editors</i>	
Taku Kudo, Toshiyuki Hanaoka, Jun Mukai, Yusuke Tabata and Hiroyuki Komatsu	19
<i>Different Input Systems for Different Devices</i>	
Asad Habib, Masakazu Iwatate, Masayuki Asahara and Yuji Matsumoto	26
<i>An Accessible Coded Input Method for Japanese Extensive Writing</i>	
Takeshi Okadome, Junya Nakajima, Sho Ito and Koh Kakusho	31
<i>Error Correcting Romaji-kana Conversion for Japanese Language Education</i>	
Seiji Kasahara, Mamoru Komachi, Masaaki Nagata and Yuji Matsumoto	38
<i>From pecher to pêcher... or pêcher: Simplifying French Input by Accent Prediction</i>	
Pallavi Choudhury, Chris Quirk and Hisami Suzuki	43
<i>Phrase Extraction for Japanese Predictive Input Method as Post-Processing</i>	
Yoh Okuno	48
<i>Robustness Analysis of Adaptive Chinese Input Methods</i>	
Mike Tian-Jian Jiang, Cheng-Wei Lee, Chad Liu, Yung-Chun Chang and Wen-Lian Hsu	53

Conference Program

Sunday November 13, 2011

8:45 Session 1: Opening and Keynote Speech

10:00 Coffee/Tea Break

10:30 Session 2: Long Papers

Challenges in Designing Input Method Editors for Indian Languages: The Role of Word-Origin and Context

Umair Z. Ahmed, Kalika Bali, Monojit Choudhury and Sowmya VB

Discriminative Method for Japanese Kana-Kanji Input Method

Hiroyuki Tokunaga, Daisuke Okanohara and Shinsuke Mori

Efficient dictionary and language model compression for input method editors

Taku Kudo, Toshiyuki Hanaoka, Jun Mukai, Yusuke Tabata and Hiroyuki Komatsu

12:00 Lunch

14:00 Session 3: Posters and Demos

Different Input Systems for Different Devices

Asad Habib, Masakazu Iwatate, Masayuki Asahara and Yuji Matsumoto

An Accessible Coded Input Method for Japanese Extensive Writing

Takeshi Okadome, Junya Nakajima, Sho Ito and Koh Kakusho

15:30 Coffee/Tea Break

16:00 Session 4: Short and Long papers

Error Correcting Romaji-kana Conversion for Japanese Language Education

Seiji Kasahara, Mamoru Komachi, Masaaki Nagata and Yuji Matsumoto

From pecher to pêcher... or pêcher: Simplifying French Input by Accent Prediction

Pallavi Choudhury, Chris Quirk and Hisami Suzuki

Sunday November 13, 2011 (continued)

Phrase Extraction for Japanese Predictive Input Method as Post-Processing

Yoh Okuno

Robustness Analysis of Adaptive Chinese Input Methods

Mike Tian-Jian Jiang, Cheng-Wei Lee, Chad Liu, Yung-Chun Chang and Wen-Lian Hsu